Introduction to rare event analysis using Monte Carlo

G. Rubino

MCQMC'18, Rennes

Outline

Outline

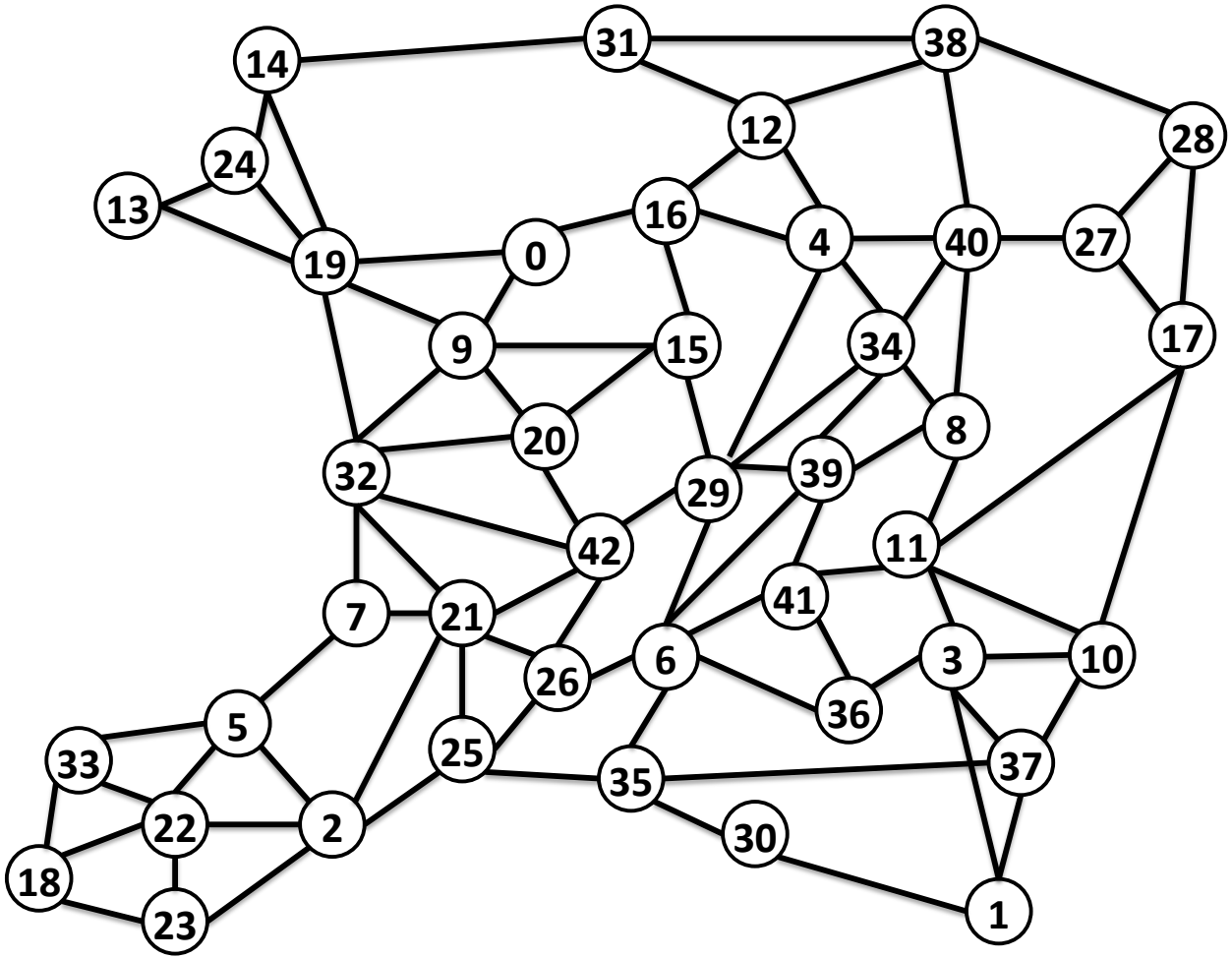
Rare events

- A rare event is an event occurring with a very small probability.
- How small? It depends on the area.
- Examples:
  - the crash of an aircraft,
  - the crash of a nuclear plant, or of a communication system,
  - a natural catastrophe,
  - etc.
- We can classify these and many other families of rare events in those with artificial or with natural causes. We will discuss only the analysis of the former; in general the tools available to analyze both types of events are different.
- To use a common illustration area, we will work here with dependability problems.

Rare events and dependability

- Dependability analysis consists in looking at systems as providers of some service, and in focusing on the fact that the service is or is not provided as specified.
- More informally, we are interested in the fact that in all real-life systems there are failures, and sometimes also repairs. These aspects of systems will be at the center of our examples.
- For instance, we don't care about the system's performance, "how much" service it provides, how fast, etc. In dependability, we focus on questions such as "is the system available for working or is it failed?", or "how much time until next system failure?", etc.
- For the analyst, the analysis takes the form of different metrics to be evaluated: the system's reliability, its availability, its Mean Time To Failure, etc.

A real-life example



European optical comm. infrastructure (43 nodes, 90 edges, 180 links)

- Call nodes the vertices of the graph, and links its arcs (we assume bi-directional communications between connected pairs of nodes).
- Assume that the 180 links can be failed or working (a binary world) (when? at some point in time t of interest, for instance at $t = \infty$,...). We say that the model is a static one.
- Say that $X_i = 1$ if link i is working, $X_i = 0$ if not, and suppose that these 180 Binary r.v.s are independent. Assume we know (after experiments inside a lab) $r_i = \mathbb{P}(X_i = 1)$ for all link i.
- An important aspect of the robustness of these topologies is the fact that, when links can fail as in previous model, it remains a path composed of working links only, between every pair of nodes, or between two important nodes, or between the nodes in some subset of the vertex set.
- Denote by R the probability of one of these previously presented connectivity events, a central parameter to compute. Call it "system's reliability", or, in words, probability that the system works. Call r_i the (elementary) reliability of link i.

- Correctly transformed into a binary output problem, computing R is, in the general case, an NP-hard problem.
- For instance, previous example in slide 6 is completely out-of-reach for any of the many algorithms available for the exact evaluation of R.
- It remains the Monte Carlo approach:
  - perform $N \gg 1$ times the following: sample the "state" of each of the 180 links, that is, sample the 180 Bernoulli variables $X_1, \ldots, X_{180}$;
  - after sampling these 180 r.v.s, check if the system works, that is, check the chosen connectivity criteria;
  - at the end, return the # of times the network worked divided by N.
- This is actually called standard or naive or crude Monte Carlo.
- For technical reasons, it is better to work with $1 - R$, the system's *unreliability*, rather than with R. We will denote $\gamma = 1 - R$.
  So, $\gamma \ll 1$.
- We also denote $u_i = 1 - r_i$, the (elementary) unreliability of link i.

Confidence intervals

- Call $Y^{(n)}$ the r.v. 1(the nth networks *fails*) in the execution of the standard Monte Carlo process.
- So, $Y^{(1)}, \ldots, Y^{(N)}$ are N independant copies of $Y \sim$ Bernoulli, with parameter γ. Recall that $\mathbb{E}(Y) = \gamma$ and $\mathbb{V}(Y) = \gamma(1-\gamma)$.
- The unknown γ is then estimated by the ratio

$$\widetilde{\gamma} = \frac{1}{N} \sum_{n=1}^{N} Y^{(n)}.$$

- Observe that $\widetilde{\gamma}$, the *standard estimator* of γ, is a r.v., with the property of being *unbiased*, which means that $\mathbb{E}(\widetilde{\gamma}) = \gamma$. Also,

$$\mathbb{V}(\widetilde{\gamma}) = \frac{1}{N^2} N \mathbb{V}(Y) = \frac{\gamma(1-\gamma)}{N}.$$

- Instead of returning simply this $\widetilde{\gamma}$, the right procedure is to compute a measure of the accuracy of the estimation, typically, a Confidence Interval for γ.
- The idea is that instead of making the computer say

  "`My estimation of` γ `is` $3.18 \cdot 10^{-9}$",

  without providing any idea about the quality of the estimation, a correct output (when using a "confidence level" of 95%, for instance), would take the form

  "`I got the confidence interval` $\left(3.04 \cdot 10^{-9}, 3.32 \cdot 10^{-9}\right)$ `for` γ`, with confidence level 0.95. The middle-point of the interval,` $3.18 \cdot 10^{-9}$`, is my point-estimation of` γ`.`"

- A standard way to do so is to apply the Central Limit Theorem which leads to the Confidence Interval

$$\left(\widetilde{\gamma} \mp c_\alpha \sqrt{\frac{\widetilde{\gamma}(1-\widetilde{\gamma})}{N-1}} \right).$$

Parameter α is a "confidence level" given beforehand, close to 1 (typical values: 0.95, 0.99, 0.999), $c_\alpha = \Phi^{-1}\big((1+\alpha)/2\big)$, where Φ^{-1} denotes the inverse of the Standard Normal c.d.f. For instance, $c_{0.95} = 1.960$, $c_{0.99} = 2.576$, $c_{0.999} = 3.291$.

- The half-size of the Confidence Interval can be interpreted as (a probabilistic bound of) the absolute error of the estimation. If we divide it by the point-estimation, we can interpret the ratio as (a probabilistic bound of) the corresponding relative error, RE.
- So, since $\gamma \approx 0$ and, then, $\widetilde{\gamma} \approx 0$, and since $N \gg 1$,

$$RE = c_\alpha \sqrt{\frac{1-\widetilde{\gamma}}{(N-1)\widetilde{\gamma}}} \approx \frac{c_\alpha}{\sqrt{N\widetilde{\gamma}}}.$$

- Given the usual values of the coefficient c_α, sometimes we remove it from previous expression, and we say that $RE = \sqrt{(1-\widetilde{\gamma})/\left[(N-1)\widetilde{\gamma}\right]} \approx 1/\sqrt{N\widetilde{\gamma}}$.
- Looking at RE better shows the problem of $\gamma \approx 0$: if we use the standard Monte Carlo approach, when the target γ is very small, so that its estimation $\widetilde{\gamma}$ is also very small, the relative error is usually very large, unless N is huge.

- Suppose that $R = 1 - 10^{-9}$, a realistic assumption in many cases (and a target in others, such as in aeronautics or telecommunications).
- Pick a classic 95% confidence for the estimations.
- Assume we want a modest 10% of relative error in the answer.
- Then, writing $1/\sqrt{N \cdot 10^{-9}} \leq 10^{-1}$ leads to $N \geq 10^{11}$. If each instance of the network needs 1 sec of CPU time for its processing (sampling and checking connectivity), we need about 12000 years to estimate γ. Too bad. And if each instance needs just 1 msec (optimistic), we still need 12 years to get the answer.
- The techniques we will mention later take usually seconds, or in some cases a few minutes, to evaluate models as the one shown before and with numbers as the preceding ones.

About typical application frameworks

- Sometimes, instead of targeting the probability of a rare event such as previous γ, we have some r.v. $Y \in \mathbb{R}^d$ for some $d \geq 1$ and a function $\psi\colon \mathbb{R}^d \rightsquigarrow \mathbb{R}$ and we look for $\mu = \mathbb{E}\big(\psi(Y)\big)$, assumed to be finite.
- In many situations, there is some Borel set A with $\mathbb{P}(A) \ll 1$, such that if we write

$$\mu = \underbrace{\int_A \psi(Y(\omega))\,\mathrm{d}\mathbb{P}(\omega)}_{a} + \underbrace{\int_{\Omega\setminus A} \psi(Y(\omega))\,\mathrm{d}\mathbb{P}(\omega)}_{b},$$

then b is very different from μ. If we estimate μ by the standard

$$\widetilde{\mu} = \frac{1}{N}\sum_{n=1}^{N} \psi(Y^{(n)}),$$

where $Y^{(1)}, \ldots, Y^{(N)}$ are N independent copies of Y, it can happen that we never, or almost never, sample in A, leading to a very poor estimation of μ, another form of the rareness problem.

- As an example in dependability, assume we model the system's dynamics by means of a stochastic process Z living in, say, $\mathbb{N}^d$ (or in some other discrete structure), with an absorbing state a. If $Z(t) \neq a$, the system is working, and if $Z(t) = a$, it is failed.
- To simplify, assume that there is a state $0 \neq a$ with $Z(0) = 0$ (the "initial" state), and that for any state $x \neq 0$ and $\neq a$, there is a path from 0 to x having probability > 0 and a path from x to state a, and not containing state 0, also with strictly positive probability.
- Think of Markov or semi-Markov continuous time processes, for instance.
- Then, a typical dependability metric is the Mean Time To Failure, or MTTF: if $T = \inf\{t > 0 \mid Z(t) = a\}$ (the *absorption time* of the process, also called in this context the system's or model's life-time), then MTTF $= \mathbb{E}(T)$.

- For any path π from 0 to a, denote by $\tau(\pi)$ its mean duration, that is: $\tau(\pi) = \mathbb{E}(T \mid \text{the process takes path } \pi)$, and by $p(\pi)$ its probability.
- Then,

$$\text{MTTF} = \sum_{\substack{\text{all paths } \pi \\ \text{from 0 to } a}} \tau(\pi) p(\pi).$$

- It often happens that there is a region in the space of the paths with a very small probability but where paths are very long (take a long time).
- This is another illustration of the problems produced by a rare event situation, this time for estimating a real number MTTF $\gg 1$.

Outline

The problem

- Consider a multi-component system composed of M independent components.
- Components and system are in a random "state" belonging to the set {up, down}. The state of component i is represented by the Binary r.v. X_i coding 1 for "up" and 0 for "down".
- The random vector $\vec{X} = (X_1, \ldots, X_M)$ is called the *system's configuration*.
- The system's state is given by the *structure function* Φ from $\{0, 1\}^M$ into $\{0, 1\}$, where $\Phi(\vec{x}) = 1$(sys. works when its configuration is $\vec{x}$). For instance, for a series of two components, $\Phi(x_1, x_2) = x_1 x_2$.
- We are given Φ and the M numbers $r_1, \ldots, r_M$ where $r_i = \mathbb{P}(X_i = 1)$. The goal is to find $R = \mathbb{P}(\Phi(\vec{X}) = 1) = \mathbb{E}(\Phi(\vec{X}))$, or equivalently, $\gamma = \mathbb{P}(\Phi(\vec{X}) = 0) = \mathbb{E}(1 - \Phi(\vec{X}))$.

- With the notation of previous section, $Y = 1 - \Phi(\vec{X})$.
- The law of $\vec{X}$ is $\pi(\vec{x}) = \prod_{i:\, x_i=1} r_i \prod_{j:\, x_j=0} (1 - r_j)$.
- The number R is the reliability of the system, and $\gamma = 1 - R$ is its unreliability. See that we have discrete sums here, but possibly with a huge number of terms:

$$R = \sum_{\vec{x} \in \{0,1\}^M} \Phi(\vec{x}) \pi(\vec{x}), \qquad \gamma = \sum_{\vec{x} \in \{0,1\}^M} \Big[1 - \Phi(\vec{x})\Big] \pi(\vec{x}).$$

- Exactly computing R or γ leads to NP-hard problems (after an appropriate transformation into a decision problem). This means here that we can not analyze exactly models even with moderate sizes.
- In the Monte Carlo case, the problem to deal with is the rare event situation, where $R \approx 1$ ($\gamma \approx 0$).

Reference problem

s

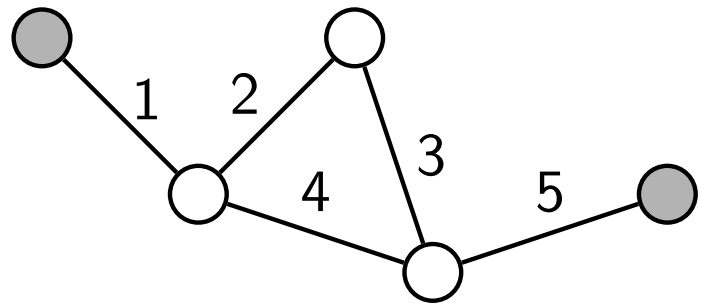
- Network reliability is a reference problem in this family (and an active research area for years).
- The main representative model is an undirected and connected graph $\mathcal{G}$ without loops, whose M edges represent the system's components (the "atomic" object that are subject to failures). In other cases, the graph is directed, or the components are the nodes, etc.
- With each configuration $\vec{x}$ we associate the partial graph $\mathcal{G}(\vec{x})$ built by removing any edge i of $\mathcal{G}$ for which $x_i = 0$.
- A subset of nodes (called *terminals*), denoted here by K, is selected, and the structure function is

$$\Phi(\vec{x}) = 1\big(\text{the nodes of } K \text{ are connected in } \mathcal{G}(\vec{x})\big).$$

- In other words, instead of giving a "table" Φ with 2^M entries, we give a graph with M edges, and Φ becomes implicitly defined.

Example 1

- For instance, let $\mathcal{G}$ be the graph

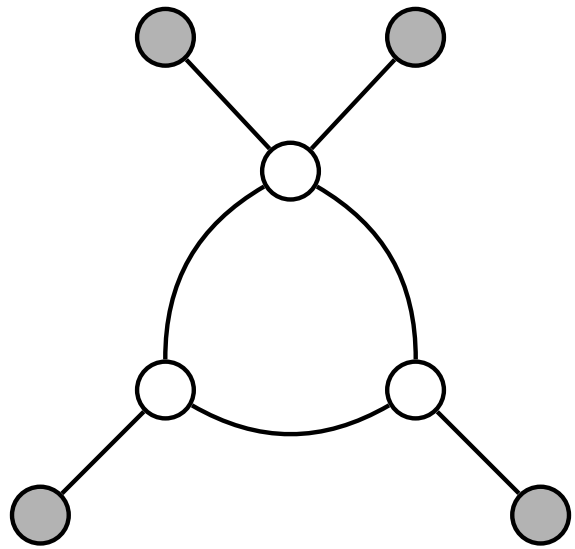


  where K is the set composed of the two grey nodes.
- We have

$$R = r_1\left[1-(1-r_4)(1-r_2r_3)\right]r_5 = r_1(r_2r_3+r_4-r_2r_3r_4)r_5$$

Example 2

Consider this other example representing a typical communication network, with K being the set of terminal machines (the grey nodes) connected through a *backbone* (the ring of white nodes).

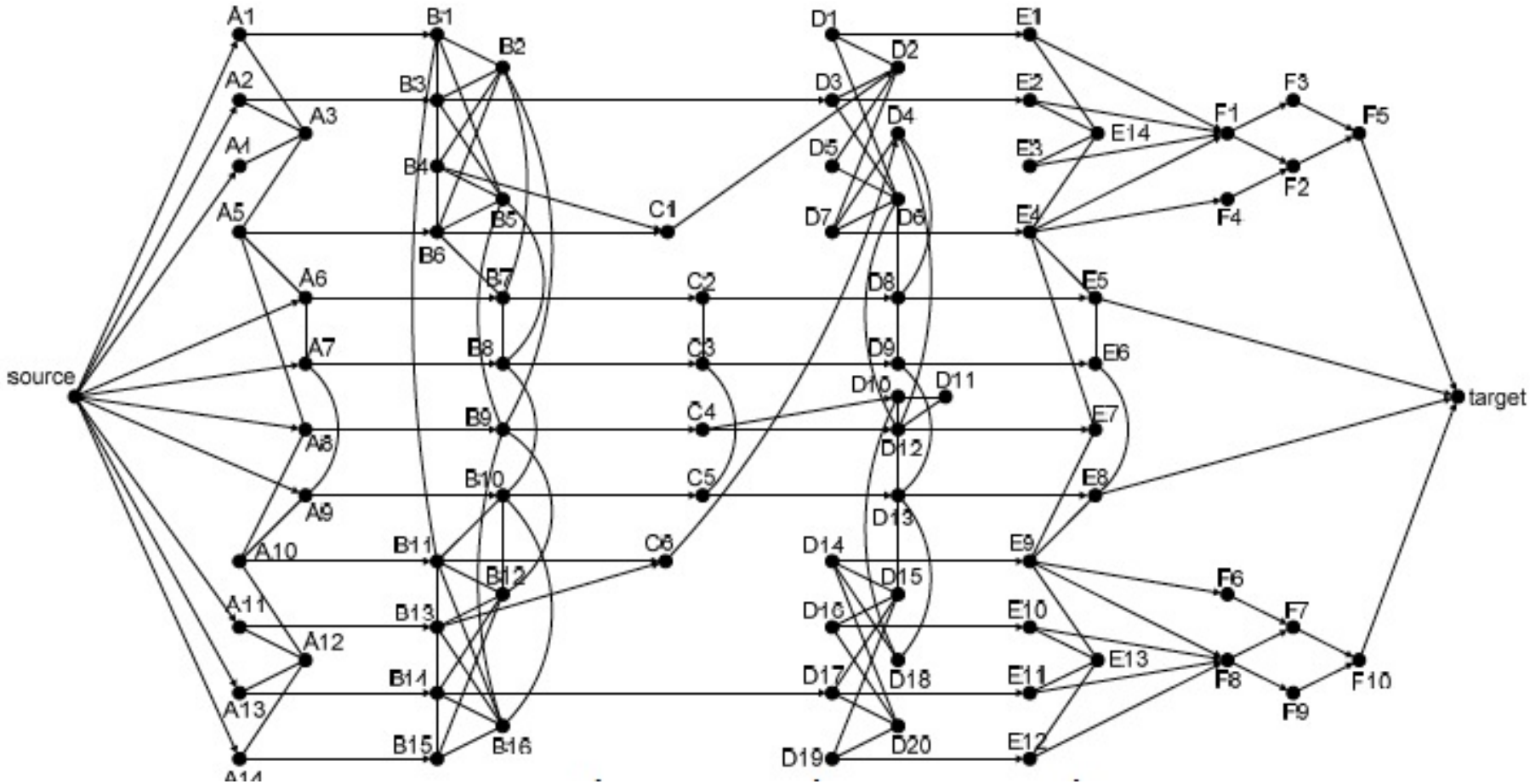


If r is the elementary reliability of any node, then

$$R = r^4\big(r^3 + 3r^2(1-r)\big) = r^6(3-2r).$$

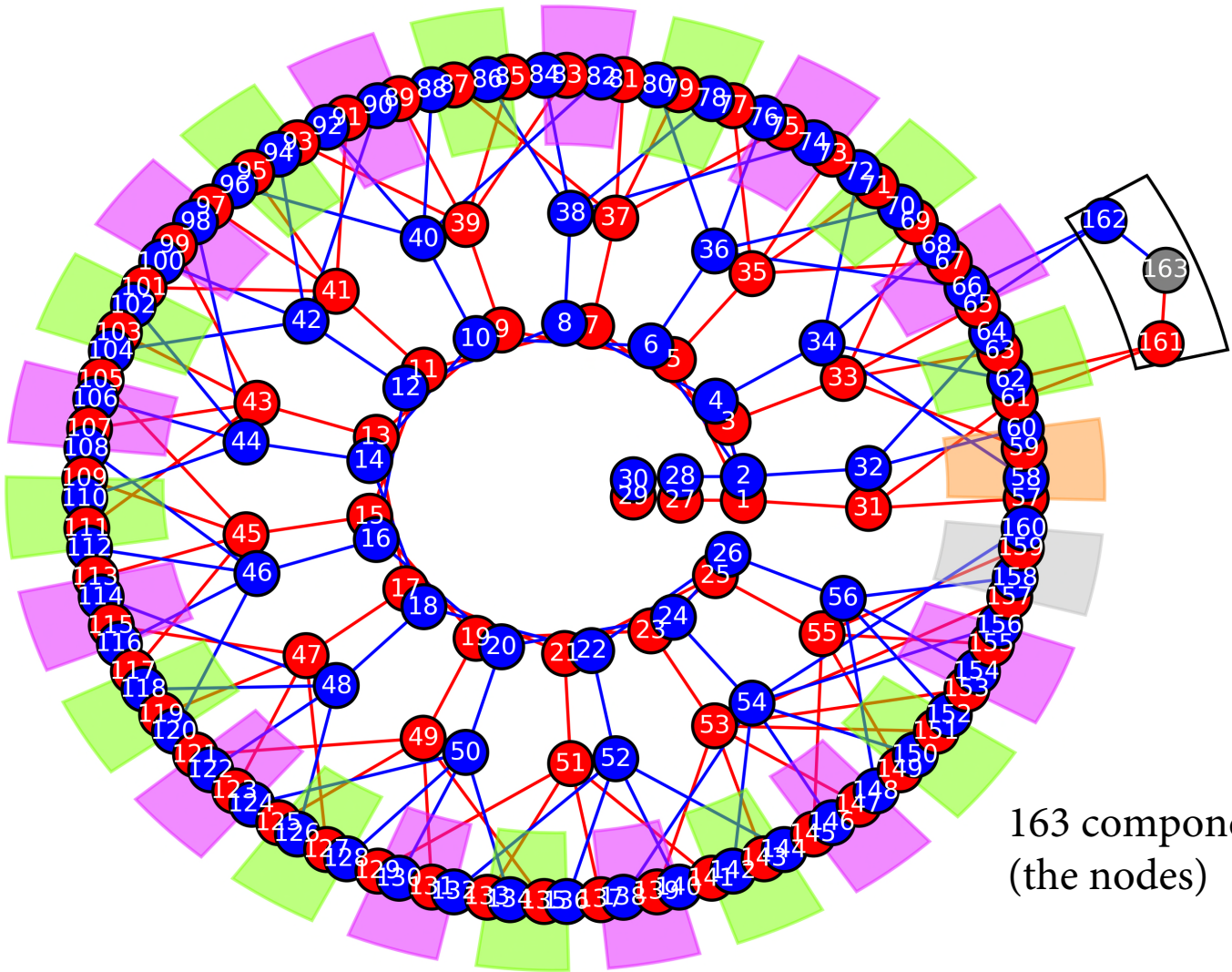
Another "out-of-reach" case

- One of many models used in the analysis of the Boeing Dreamliner 787 aircraft:



- 82 nodes, 171 links

An example from a subway model



163 components
(the nodes)

- This example of a specific communication system (and many other similar ones) is currently under analysis by our team at INRIA.
- It comes from communication systems developed by ALSTOM for its last underground train systems.

Outline

Specialized techniques

The rarity aspect precludes the use of a naive (classical, standard, direct) approach: specialized simulation techniques are then needed.

The specialized simulation techniques designed for rare event analysis belong usually to some well studied family of methods, such as

- **Importance Sampling techniques**
  (with, in particular, Zero-Variance versions)
- **Splitting techniques** (mainly for dynamic models)
- Recursive Variance Reduction techniques (mainly for static models)
- ...

Alternatives

- **Bounding techniques**. In some rare event situations, we can build very tight bounds of the metrics of interest. More than that, in several cases of strong practical interest, the bounds improve when the event becomes rarer.
- **Mean field analysis**. Powerful and very useful approach in specific families of models.
- ...

Splitting in a nutshell

- Designed for dynamic models.
- Imagine we need to estimate the probability of a collision in a transportation network, for instance, between vehicles in a city, under normal traffic conditions, not in a heavy traffic situation.
- For instance, we want to evaluate how rare this event is, and to be more specific, suppose we want to analyze collisions occurring when long queues have formed.
- If the event is rare, we would have to wait probably too much until queues start to form during the simulation, because we don't want to consider heavy traffic.
- In Splitting, we define, for instance, a set of quantitative queue levels, say $L_1, L_2, \ldots, L_H$, with $L_1 < L_2 < \cdots < L_H$ (for instance, levels can be measured by a number of vehicles).

- When a queue starts to form and its size, say, reaches value L_1, we stop the sim, and from that point, we start K_1 new independent simulations and wait until next level (or threshold) L_2 is reached in some of the K_1 independent or parallel "trajectories".
- If in a trajectory the queue reaches level L_2, then we split it again in many parallel and independent paths, say K_2 of them, and the process continues.
- If one of those runs comes back to a previous level, or in other variants to a state where the queue has disappeared, one option is to kill it (to avoid wasting computing resources).
- We can prove that if one or several runs reach last level, we can start from that state independent runs where collision is now a much more frequent event, evaluate its probability, and deduce an unbiased estimation of the collision probability we were looking for at the beginning.

- We can feel, at least intuitively, that if the levels are close to each other and if the splitting factors K_i are high, we increase the probability of reaching high levels but we increase also the cost.
- This suggests that there are optimal settings and that we can expect an increase in efficiency with respect to a naive simulation procedure.
- This is actually what happens in practice, but, in general, after specific design efforts that depend on the problem at hand, and that may fail in other contexts.
- So, as usual in the field, there is no universal receipt, but guidelines instead supported by some general results valid on simplified models.
- On the positive side, gains can be in several orders of magnitude in variance, and also in global efficiency (that is, when taking into account both the variance and the execution time).

Importance Sampling

- To put things as simple as possible, assume we have some random vector $\vec{X} \in \mathbb{N}^d$ and we need to estimate $\gamma = \mathbb{P}(\vec{X} \in A) \ll 1$.
- Let $p()$ denote the probability mass function of $\vec{X}$. We have

$$\gamma = \sum_{\vec{x} \in \mathbb{N}^d} 1(\vec{x} \in A) p(\vec{x}).$$

- In many cases, $p()$ is unknown, or it is known but the sum has a huge number of terms, but it is however possible to efficiently sample from $p()$, so that we can estimate γ by simulation.

- Now, assume that $\widehat{p}(\,)$ is another mass probability function on $\mathbb{N}^d$, such that if $\vec{x} \in A$ and $p(\vec{x}) > 0$, then $\widehat{p}(\vec{x}) > 0$ as well. We have

$$\gamma = \sum_{\vec{x}\in\mathbb{N}^d} 1(\vec{x} \in A) p(\vec{x}) = \sum_{\vec{x}\in\mathbb{N}^d:\, \vec{x}\in A,\, p(\vec{x})>0} p(\vec{x}) = \sum_{\vec{x}\in\mathbb{N}^d:\, \vec{x}\in A,\, p(\vec{x})>0} \frac{p(\vec{x})}{\widehat{p}(\vec{x})} \widehat{p}(\vec{x}).$$

- Then, denoting $L(\vec{x}) = p(\vec{x})/\widehat{p}(\vec{x})$ when $\widehat{p}(\vec{x}) > 0$, and 0 otherwise (called the *likelihood* ratio), $\gamma = \sum_{\vec{x}\in\mathbb{N}^d} 1\big(\vec{x} \in A\big) L(\vec{x}) \widehat{p}(\vec{x})$ can be written

$$\gamma = \mathbb{E}_{\widehat{p}}(1\big(\vec{Y} \in A\big) L(\vec{Y})).$$

- This means that a new estimator of γ is

$$\widehat{\gamma} = \frac{1}{N} \sum_{n=1}^{N} 1\big(\vec{Y}^{(n)} \in A\big) L(\vec{Y}^{(n)}),$$

where $\vec{Y}^{(1)}, \ldots, \vec{Y}^{(N)}$ are N independent copies of $\vec{Y} \sim \widehat{p}$.

- The corresponding Confidence Interval is

$$\left(\widehat{\gamma} \mp c_\alpha \frac{\widehat{\sigma}}{N}\right),$$

where

$$\widehat{\sigma}^2 = \frac{1}{N-1} \sum_{n=1}^{N} 1\big(\vec{Y}^{(n)} \in A\big) L(\vec{Y}^{(n)}) - \frac{N}{N-1} \widehat{\gamma}^2.$$

- So, we are estimating γ by estimating an artificial new thing (the expectation of the r.v. $1\big(\vec{Y} \in A\big)L(\vec{Y})$ where $\vec{Y} \in \mathbb{N}^d$ follows the new dynamics $\widehat{p}(\,)$).
- We say that we did a "change of measure" (c.o.m.), from $p(\,)$ to $\widehat{p}(\,)$.
- The question is how good is this new estimator $\widetilde{\gamma}$ (that is, how small is its variance and how expensive is sampling from $\widehat{p}(\,)$?
- The answer is that by choosing $\widehat{p}(\,)$ carefully, we may obtain estimators thousands of times more efficient than the standard one. More precisely, with the same cost (execution time), which often reduces basically to take the same sample size N, the variance of the IS estimator can be thousands of times smaller than the variance of the standard one.

Comparing variances

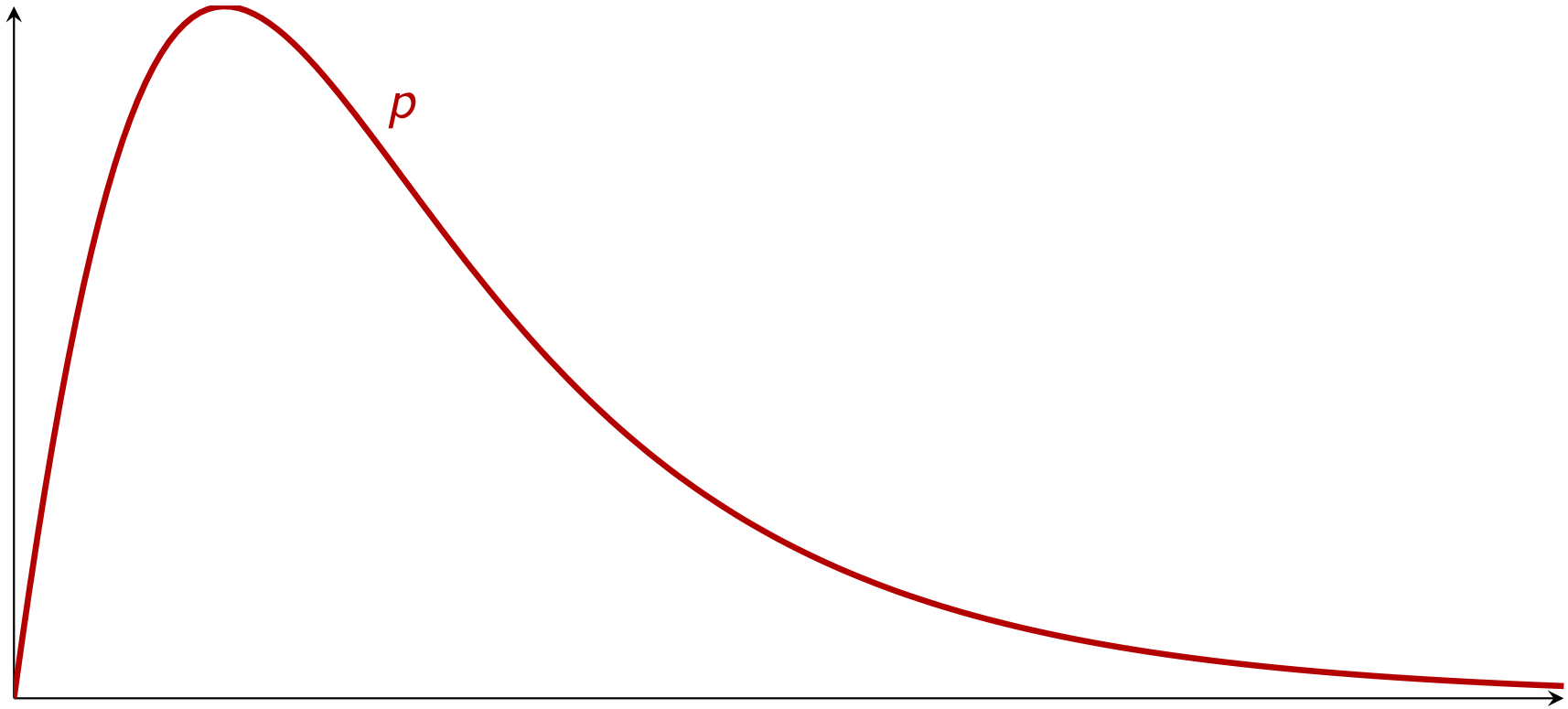
- Let us compare the variances of the standard estimator of γ, denoted $\widetilde{\gamma}$, and of the IS estimator $\widehat{\gamma}$, for the same sample size N.
- So, $\widetilde{\gamma} = N^{-1} \sum_{n=1}^{N} 1\big(\vec{X}^{(n)} \in A\big)$ where $\vec{X}^{(1)}, \ldots, \vec{X}^{(N)}$ are independent copies of $\vec{X} \sim p(\,)$, and $\widehat{\gamma} = N^{-1} \sum_{n=1}^{N} 1\big(\vec{Y}^{(n)} \in A\big) L(\vec{Y}^{(n)})$ where $\vec{Y}^{(1)}, \ldots, \vec{Y}^{(N)}$ are independent copies of $\vec{Y} \sim \widehat{p}(\,)$.
- See that

$$\mathbb{V}(\widetilde{\gamma}) = \frac{1}{N} \mathbb{V}\big(1(\vec{X} \in A)\big) \qquad \text{and} \quad \mathbb{V}(\widehat{\gamma}) = \frac{1}{N} \mathbb{V}\big(1(\vec{Y} \in A) L(\vec{Y})\big).$$

- Writing $\mathbb{V}(\widetilde{\gamma}) = N^{-1}\left[\mathbb{E}\big(1(\vec{X} \in A)^2\big) - \gamma^2\right]$ and $\mathbb{V}(\widehat{\gamma}) = N^{-1}\left[\mathbb{E}\big(1(\vec{Y} \in A)^2 L^2(Y)\big) - \gamma^2\right]$, comparing variances reduces to comparing second moments.
- So, IS is better than the standard estimator iff $\mathbb{E}\big(1(\vec{X} \in A)\big) > \mathbb{E}\big(1(\vec{Y} \in A) L^2(Y)\big)$.
- We see then that the element that controls this issue is the value of the likelihood ratio, and what we need is $\widehat{p}(\vec{x}) > p(\vec{x})$ when we are in A (and when $p(\vec{x}) > 0$).

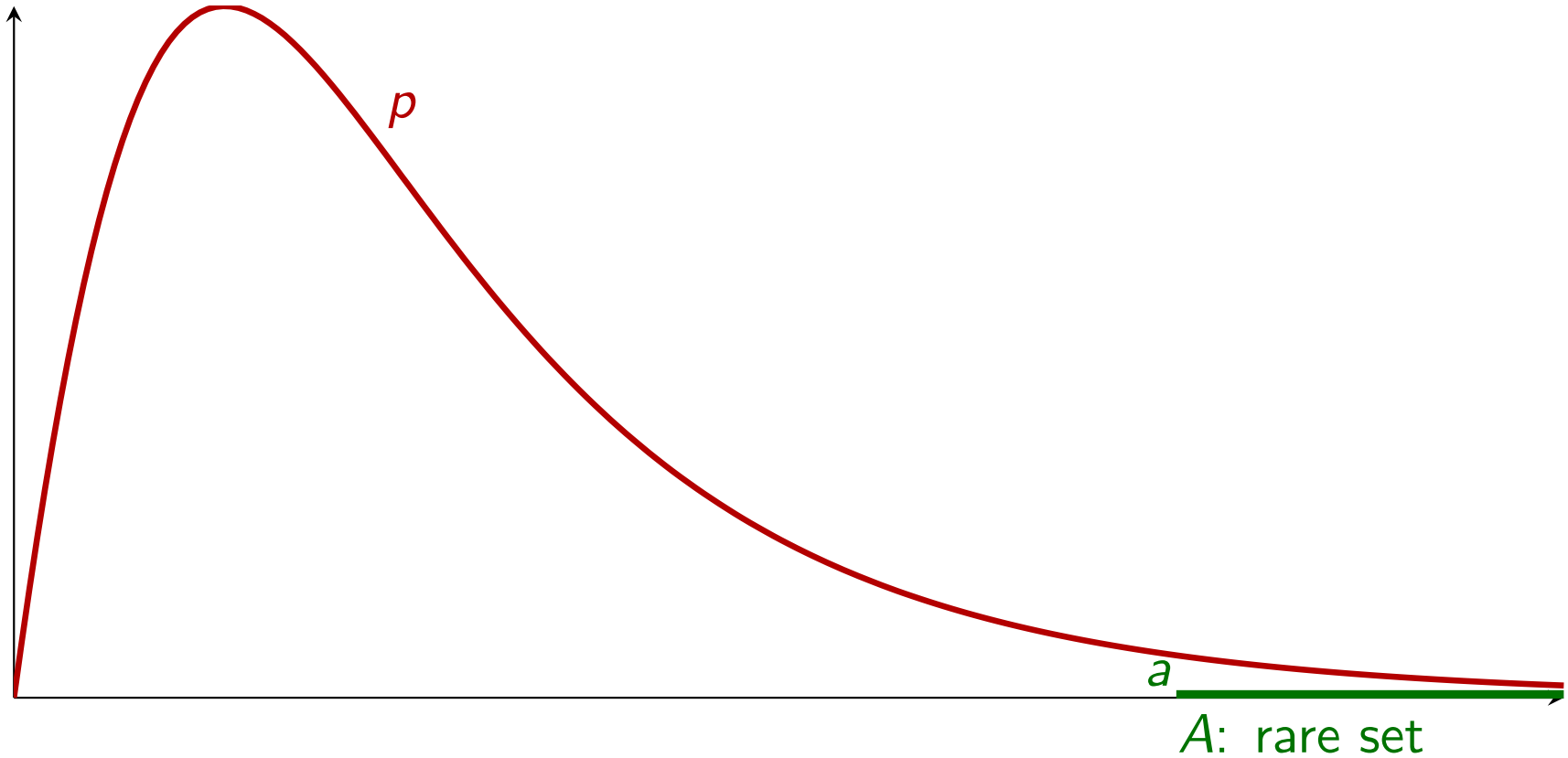
IS illustration

The original law of X (seen "from so far" that it looks continuous).



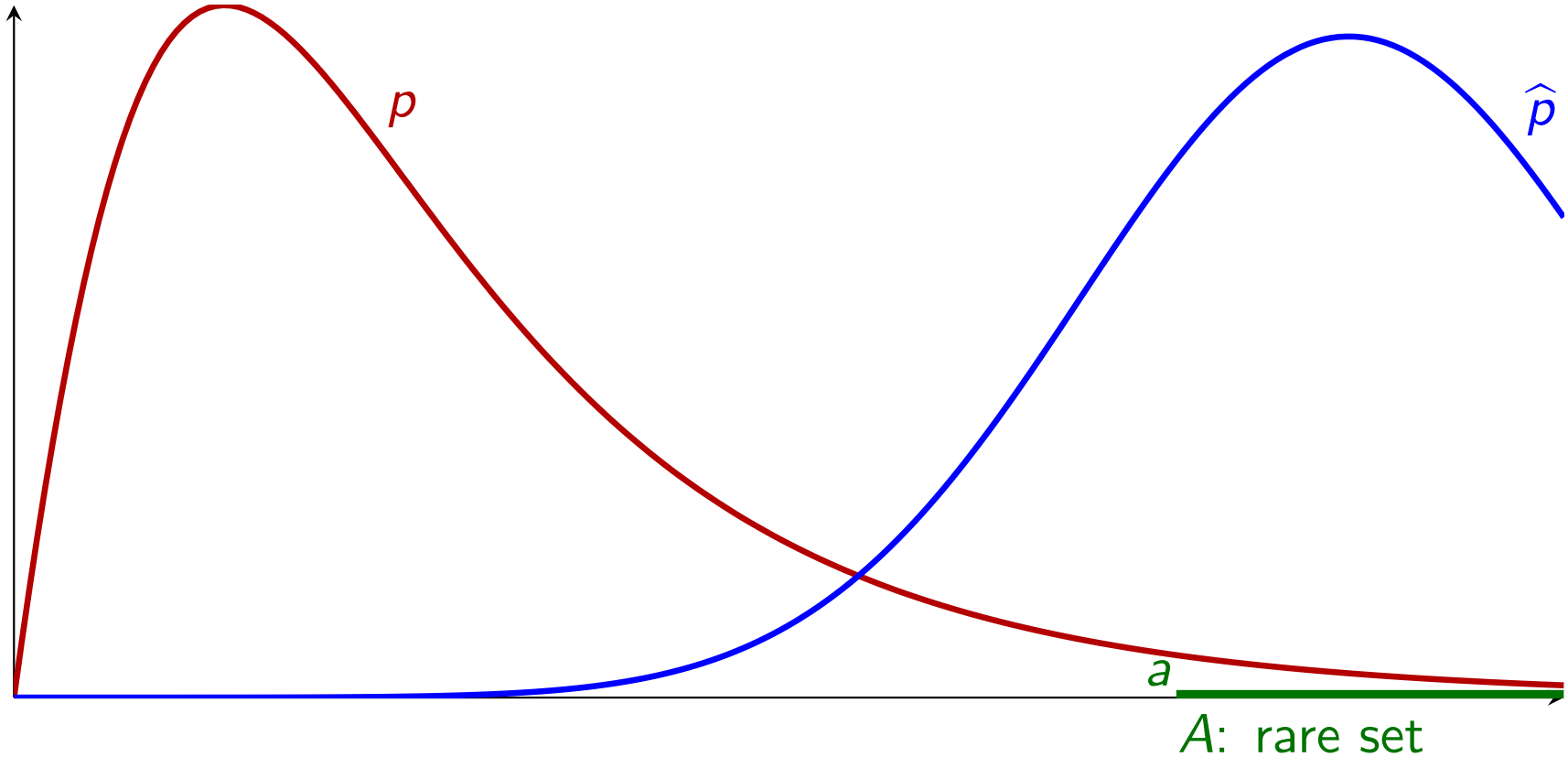
IS illustration

The rare set A.



IS illustration

The change of measure $p \longrightarrow \widehat{p}$.



On the perfect IS estimator

- Consider the c.o.m.

$$\widehat{p}(\vec{x}) = \frac{1(\vec{x} \in A)p(\vec{x})}{\gamma}.$$

- We have

$$\mathbb{V}(\widehat{\gamma}) = \mathbb{V}\big(1(\vec{Y} \in A)L(\vec{Y})\big) = 0.$$

- You "sample" only once ($N = 1$), and the result is exactly γ.
- Of course, this looks useless since we need the target (the value of γ) to implement it. More on this later.
- This suggests as a general rule, to try to make the new dynamics proportional to the numerator, something that "follows" the shape of $p()$ on A.
- A more "high level" rule of thumb is "make rare things frequent" in the new dynamics ... "but not too much".

Making rare things too frequent

A classic toy but illustrative example.

- Suppose that the life-time L of a system follows the Exponential distribution[1] with parameter λ. In such a context, λ is the system's *failure rate*. We want to estimate the probability γ that the failure arrives before a given time T.
- We assume that we know λ (in which case, of course we know the answer:

$$\gamma = \mathbb{P}(L \leq T) = 1 - \mathrm{e}^{-\lambda T}.$$

- We will estimate γ using the standard estimator, and then, we will try IS using as the new measure another Exponential law with some rate $\widehat{\lambda}$ that we will vary, to explore the efficiency of IS as a function of it.

[1] This is reasonable for many types of systems.

- Write $\psi(t) = 1(t \leq T)$. The target is

$$\gamma = \int_0^\infty \psi(x) f(x)\,\mathrm{d}x = \int_0^T \lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x = 1 - \mathrm{e}^{-\lambda T}.$$

- If $t \geq 0$, the likelihood ratio at t is

$$L(t) = \frac{f(t)}{\widehat{f}(t)} = \frac{\lambda\,\mathrm{e}^{-\lambda t}}{\widehat{\lambda}\,\mathrm{e}^{-\widehat{\lambda} t}}.$$

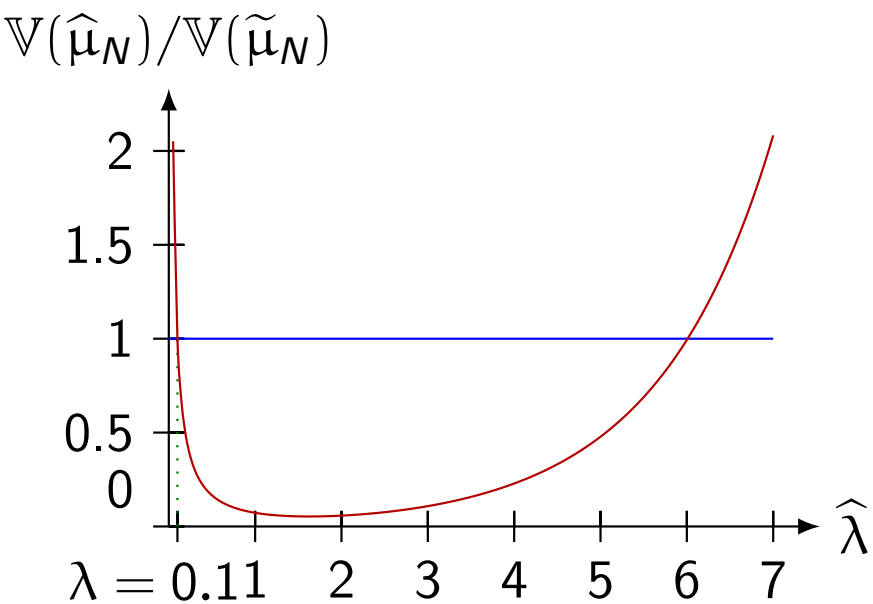
- If we compare the second moments, we have that the IS method is better (smaller variance) than the standard one if and only if

$$\frac{\int_0^\infty \psi^2(t) L(t) f(t)\,\mathrm{d}t}{\int_0^\infty \psi^2(t) f(t)\,\mathrm{d}t} = \frac{\int_0^T L(t) f(t)\,\mathrm{d}t}{\int_0^T f(t)\,\mathrm{d}t} \leq 1.$$

- For instance,

$$\int_0^T L(t) f(t)\,\mathrm{d}t = \int_0^T \frac{\lambda\,\mathrm{e}^{-\lambda t}}{\widehat{\lambda}\,\mathrm{e}^{-\widehat{\lambda} t}} \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = \frac{\lambda^2}{\widehat{\lambda}} \frac{1-\mathrm{e}^{-(2\lambda-\widehat{\lambda})T}}{2\lambda-\widehat{\lambda}}.$$

- Once we do the computations for the variance of the IS estimator (messy), we can analyze the evolution of the ratio $\mathbb{V}(\widehat{\gamma})/\mathbb{V}(\widetilde{\gamma})$ with $\widehat{\lambda}$. See here a graphical view when we fix $\lambda = 1$ and $T = 0.1$:



- The point is that if we start with a small value of $\widehat{\lambda}$, when we increase it the IS method improves in efficiency, and when $\widehat{\lambda}$ becomes greater than some $\widehat{\lambda}_0$, the IS estimator has a smaller variance than the standard one.
- But after a second threshold $\widehat{\lambda}_1$, the IS estimator becomes worse than the standard one, and its variance grows without any bound as $\widehat{\lambda} \uparrow \infty$.
- This simple example shows that we must make the rare event become frequent under the new measure, *but not excessively frequent*.
- We can anticipate then the fact that finding a good change of measure is in many cases hard, and there is no general way to do it. It strongly depends on the problem.

Best c.o.m. inside a specific family

- Remember the design rule of thumb: "use a change of measure that makes frequent what was rare, but moderately".
- In the static world, a first and simple idea is straightforward: change the individual unreliabilities (which are close to 0) into numbers that are, say, between 0.5 and 0.8. This is called "Failure Biasing" (FB) and it already works fine (even if it is possible to do much better).
- As a particular but frequent case, assume an homogeneous situation where the components' reliabilities are all the same, $= r$, and we want to use FB by changing r into some $r' < r$.
- We can prove that the optimal value r' (optimal inside this specific family of FB c.o.m.) is $r' = c/M$, where M is the number of links and c is the size of a cut of minimal size.

- As a more generic example, there is a methodology called cross entropy that can be used to find the best c.o.m. inside a given parametric family of distributions. See the last session on rare event analysis next Friday morning.
- It is possible to design a *sequential* way of simulating the system using IS, and modify the c.o.m. as a function of the evolution of the simulation.
- Previous way to operate is called state-dependent or adaptive IS. We typically find it when the model is a stochastic process, but it can appear when dealing with static models. The idea is to build an artificial stochastic process related to the static model (the "sequential way" mentioned above) and then, to adapt the c.o.m. according to observed events appearing during the simulation process.

Back at the optimal c.o.m.

- Recently, the optimal and in principle useless change of measure led to a very powerful idea, that became a hot topic these days in the area: the zero-variance approach.
- The idea is to transform this global view "$\widehat{p}(\vec{x}) = 1(\vec{x} \in A)p(\vec{x})/\gamma$" into "local relationships inside the model", local connection between different quantities related to the many components of the model.
- As we could expect, the global form of the optimal $\widehat{p}$ that requires to know the exact answer to the original question (so, no need to simulate) translates into "atomic" or "detailed" relations where things like γ appear.
- The gain happens when instead of $\widehat{p}$ we use some $\widehat{p}^*$, built using these detailed expressions, where we replace the forever unknown exact quantities by approximations, even rough ones. The resulting IS procedure is sometimes extremely efficient.

In MCQMC'18

Sessions on rare event problems and/or specialized families of Monte Carlo methods:

- Monday 16:15 - 18:15, room 5 (page 88)
- Tuesday 16:15 - 17:45, Amphi 3 (page 129)
- Wednesday 10:00 - 12:00, room 7 (page 100)
- Wednesday 13:45 - 15:45, room 8 (page 106)
- Friday 10:00 - 12:00, room 6 (page 110)

Zoom

- In the rest of the tutorial, I will focus on the main ideas of our 2015 paper "*Balanced and Approximate Zero-Variance Recursive Estimators for the Static Communication Network Reliability Problem*", TOMACS, 25(1): 5:1-5:19, 2015, co-authored with H. Cancela, M. El Khadiri and B. Tuffin.
- It describes of a new approach combining a powerful recursive estimation technique with an Importance Sampling scheme approximating the zero-variance one.
- We claim that it is (at least) at the top positions in the state-of-the-art technology for solving these rare events problems associated with a network reliability problem.
- The paper allows also to illustrate important properties of the estimators, as well as the problems of proving them on specific cases.

Outline

Bounded Relative Error

- An unbiased estimator Y' of $\mathbb{E}[Y]$ has *Bounded Relative Error* (BRE) if *RE* remains bounded as the event becomes rarer.
- Formally, we have BRE if $\sqrt{\mathbb{V}(Y')}/\mathbb{E}[Y]$ seen as a family of functions indexed by the sample size N, is uniformly bounded when $\mathbb{E}[Y] \to 0$, or equivalently, if $\mathbb{E}\left(Y'^2\right)/\mathbb{E}^2(Y)$ is uniformly bounded as $\mathbb{E}[Y] \to 0$.
- BRE implies that the sample size required to get a given relative error is not sensitive to the rarity of the event.
- The standard estimator does not possess this property.

Other relevant properties of estimators

- **Weaker than BRE**. An unbiased estimator Y' of $\mathbb{E}[Y]$ is *Asymptotically Optimal* (AO) or *Logarithmically Efficient* (LE) if

$$\lim_{\mathbb{E}(Y) \to 0} \frac{\ln \mathbb{E}\left(Y'^2\right)}{\ln \mathbb{E}(Y)} = 2.$$

- **Stronger than BRE**. An unbiased estimator Y' of $\mathbb{E}[Y]$ verifies the *Vanishing Relative Error* (VRE) property if $\sqrt{\mathbb{V}(Y')}/\mathbb{E}[Y] \to 0$ as $\mathbb{E}[Y] \to 0$.

On the AO or LE property

- To understand the connection of the definition of AO with the subject, consider the IS estimator $\widehat{\gamma}$ of previous section and recall the expression

$$\mathbb{V}(\widehat{\gamma}) = \frac{1}{N}\left[\mathbb{E}_{\widehat{p}}\big(1(\vec{Y} \in A)^2 L^2(Y)\big) - \gamma^2\right].$$

- Denote $E = \mathbb{E}_{\widehat{p}}\big[1\big(\vec{Y} \in A\big)L^2(Y)\big]$. We have $\mathbb{V}(\widehat{\gamma}) = (E - \gamma^2)/N$, so, $E = N\mathbb{V}(\widehat{\gamma}) + \gamma^2 \geq \gamma^2$.
- Then, $\ln E \geq 2\ln\gamma$, and

$$\frac{\ln E}{\ln \gamma} \leq 2.$$

- See that $\ln E = 2\ln\gamma$ leads to $\mathbb{V}(\widehat{\gamma}) = 0$.

Rarity parameter

- More generally, we often use a *rarity parameter* to "move the model toward a rare situation" and explore the behavior of estimators (a kind of asymptotic analysis).
- For instance, in the static case, we can set
  - $r_i = r = 1 - \varepsilon$ (homogeneous model), and let $\varepsilon \longrightarrow 0$,
  - or $r_i = 1 - a_i \varepsilon^{b_i}$ (heterogeneous model), with $a_i > 0$ and $b_> 0$, and let $\varepsilon \longrightarrow 0$,

  and then define BRE, AO, VRE, etc., with respect to $\varepsilon \longrightarrow 0$.
- If the model is a queuing one and the rare event is the saturation of the queue whose storage capacity is H, then the rarity parameter can be H when $H \uparrow \infty$.

Other measures

- There is another set of properties where we look not only at the variance but also to the cost in time of the execution of the algorithm implementing the estimator.
- Let us denote by $\tau(\widehat{\gamma}_N)$ the mean execution time of the estimation code where we added explicitly the sample size N.
- Then what we want is to have $\mathbb{V}(\widehat{\gamma}_N)$ small and also $\tau(\widehat{\gamma}_N)$ small. This is the basis of another set of definitions. For instance, the relative efficiency of the estimator is defined as the ratio $\gamma/\big(\mathbb{V}(\widehat{\gamma}_N)\tau(\widehat{\gamma}_N)\big)$.
- Then, we define Bounded Relative Efficiency, etc.

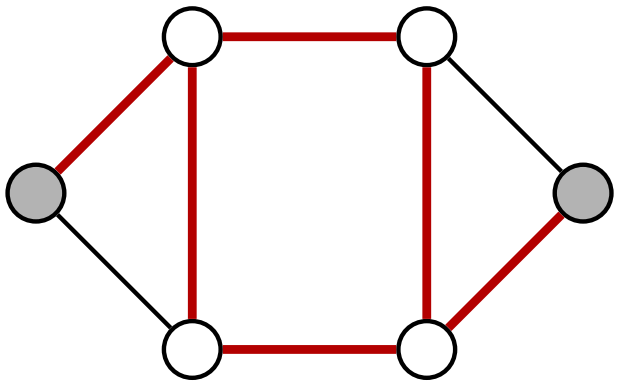
Outline

Notation and terminology

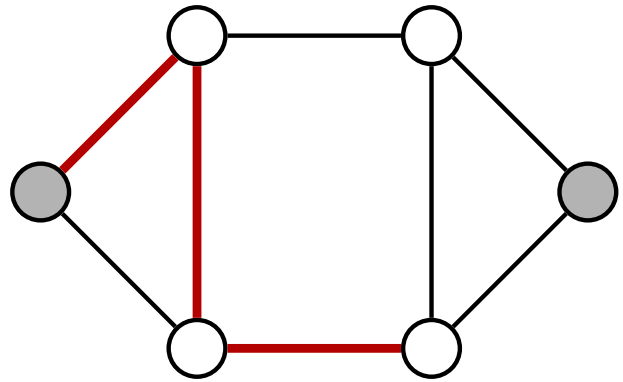
- We will keep the notation R and we will denote $Q = 1 - R$. In the same way, we will use $q_i = 1 - r_i$ for any component i (this is just to get closer to the notation used in the paper we are following).
- Cuts in a structure function (or in a graph setting, to simplify):
  - a *cut* is a set of components such that if they are all down, the system is down;
  - a *mincut* is a cut that has no strict subset that is also a cut.

Examples of cuts

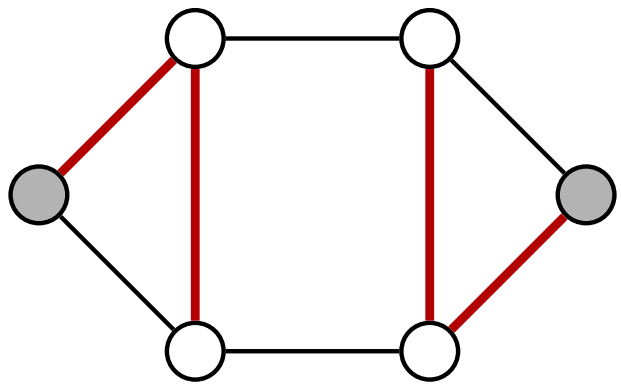
A cut but not a mincut:



A mincut:

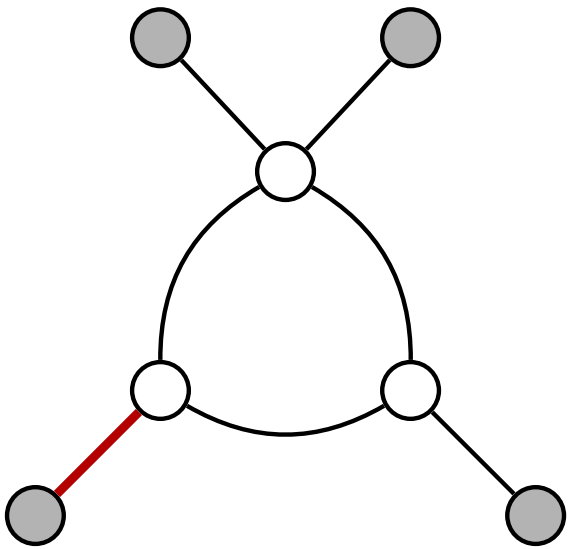


Another mincut:

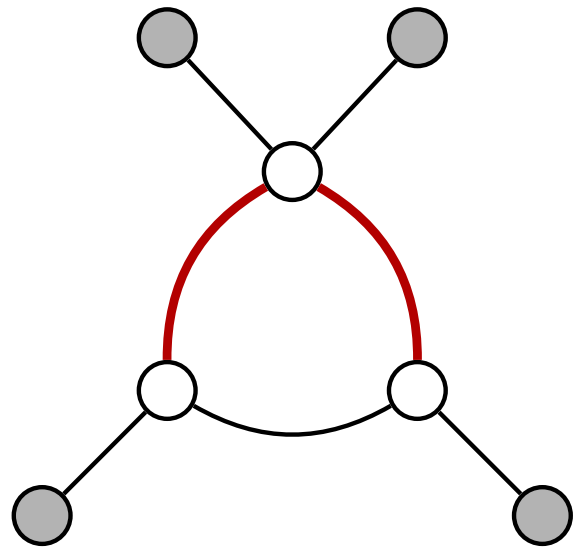


More examples of cuts

A mincut:



Another mincut:



Recursive Variance Reduction (RVR)

- Principle: choose a cutset, i.e., a set $\mathcal{C}$ of links whose failure ensures the system failure.
- Denote by $q_{\mathcal{C}}$ the probability that all links in $\mathcal{C}$ are failed, that is, $q_{\mathcal{C}} = \prod_{i \in \mathcal{C}} q_i$. By notation consistency, the target Q can be also written $q(\mathcal{G})$.
- By definition, if all links in $\mathcal{C}$ are failed, the system is failed. Consequently, $q_{\mathcal{C}} \leq Q$.
- Put some order on the links of $\mathcal{C}$. Let's denote $\mathcal{C} = \{1, 2, \ldots\}$.
- For $j = 1, 2, \ldots$, let $B_j =$ "the $j-1$ first links of $\mathcal{C}$ are down, but the jth is up".
- $\mathbb{P}[B_j] = \left(\prod_{k=1}^{j-1} q_k\right) r_j$.
- Define the conditional probabilities $p_j = \mathbb{P}[B_j \mid \text{at least one link of } \mathcal{C} \text{ is working}] = \mathbb{P}[B_j]/(1 - q_{\mathcal{C}})$.

Recursive Variance Reduction (RVR)

The RVR estimator:

- Choose a cutset, and compute $q_{\mathcal{C}}$ and the p_js.
- Pick an edge at random in $\mathcal{C}$ according to the probability distribution $(p_j)_{j=1,\cdots,|\mathcal{C}|}$.
- Let the chosen edge be the jth. Call $\mathcal{G}_j$ the graph obtained from $\mathcal{G}$ by deleting the first $j-1$ edges of $\mathcal{C}$ and by contracting the jth.
- The value y_{RVR} returned by the RVR estimator of $q(\mathcal{G})$, the estimation of the unreliability of $\mathcal{G}$, is recursively defined by

$$y_{RVR}(\mathcal{G}) = q_{\mathcal{C}} + (1 - q_{\mathcal{C}}) y_{RVR}(\mathcal{G}_j).$$

RVR estimator

Formally, the RVR estimator of $Q = q(\mathcal{G})$ is the random variable

$$Y_{RVR} = q_{\mathcal{C}} + (1 - q_{\mathcal{C}}) \sum_{j=1}^{|\mathcal{C}|} \frac{\mathbf{1}_{B_j}}{1 - q_{\mathcal{C}}} Y_{RVR}(\mathcal{G}_j).$$

Theorem

The estimator is unbiased: $\mathbb{E}[Y_{RVR}] = q(\mathcal{G}) = Q$.

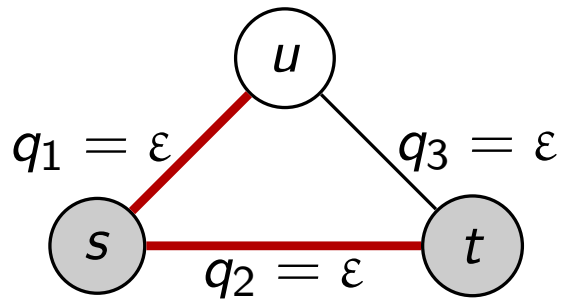
Proof: induction using the recursion.

Second moment is

$$\begin{aligned}\mathbb{E}[Y_{RVR}^2] &= q_{\mathcal{C}}^2 + 2q_{\mathcal{C}}(1 - q_{\mathcal{C}}) \left(\sum_{j=1}^{|\mathcal{C}|} \frac{\mathbb{P}[B_j]}{1 - q_{\mathcal{C}}} \mathbb{E}[Y_{RVR}(\mathcal{G}_j)] \right) \\ &\quad + (1 - q_{\mathcal{C}})^2 \left(\sum_{j=1}^{|\mathcal{C}|} \frac{\mathbb{P}[B_j]}{1 - q_{\mathcal{C}}} \mathbb{E}[Y_{RVR}^2(\mathcal{G}_j)] \right).\end{aligned}$$

No Bounded Relative Error for RVR

The RVR algorithm does not verify the Bounded Relative Error property in general. Consider the example

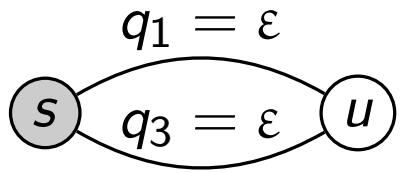


- Selected cut: the two red links, ordering them as first the link from s to t.
- $q_{\mathcal{C}} = \varepsilon^2$.

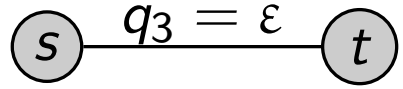
$$\begin{aligned}\mathbb{E}[Y_{RVR}^2] &= \varepsilon^4 + 2\varepsilon^2\Big[(1-\varepsilon)\mathbb{E}[Y_{RVR}(\mathcal{G}_1)] + \varepsilon(1-\varepsilon)\mathbb{E}[Y_{RVR}(\mathcal{G}_2)]\Big] \\ &+ (1-\varepsilon^2)\Big[(1-\varepsilon)\mathbb{E}[Y_{RVR}^2(\mathcal{G}_1)] + \varepsilon(1-\varepsilon)\mathbb{E}[Y_{RVR}^2(\mathcal{G}_2)]\Big].\end{aligned}$$

where

- $\mathcal{G}_1$: link from s to t is working $\rightsquigarrow$ s and t are merged (the system is necessarily connected).



  $Y_{RVR}(\mathcal{G}_1) = 0$. Thus $\mathbb{E}[Y_{RVR}(\mathcal{G}_1)] = \mathbb{E}[Y^2_{RVR}(\mathcal{G}_1)] = 0$.
- $\mathcal{G}_2$: link from s to t failed, but the one from s to u is working $\rightsquigarrow$ s and u are merged.



  $\mathbb{E}[Y_{RVR}(\mathcal{G}_2)] = \varepsilon$, $\mathbb{E}[Y^2_{RVR}(\mathcal{G}_2)] = \varepsilon^2$.
- Finally, $\mathbb{E}[Y^2_{RVR}] = \Theta(\varepsilon^3)$, and $\mathbb{E}[Y^2_{RVR}]/(\mathbb{E}[Y_{RVR}])^2 = \Theta(\varepsilon^{-1}) \to \infty$ as $\varepsilon \to 0$.

Observe that on a specific graph we may have BRE or not, depending on the ordering of the links.

Outline

Balanced RVR

- Non-BRE comes from the use of the "crude distribution" when sampling the first working link on the cut.
- Idea: use Importance Sampling (IS) instead; that is, the sampling distribution of the first line up in the cut, knowing that there is at least one, is not anymore (p_j).
- So far, we built a partition by assigning to the events B_j, for $1 \leq j \leq |\mathcal{C}|$, the conditional probabilities

$$p_j = \mathbb{P}\big[B_j \mid \text{at least one link of } \mathcal{C} \text{ is working}\big].$$

- Let us write the RVR estimator as

$$Y_{RVR} = q_{\mathcal{C}} + (1 - q_{\mathcal{C}}) \sum_{j=1}^{|\mathcal{C}|} \mathbf{1}_{B_j'} Y_{RVR}(\mathcal{G}_j),$$

where B_j' represents the same event as B_j but with the (conditional) probability p_j.

Balanced RVR (BRD)

- Now, we change the probability p_j by the uniform distribution on $\{1, 2, \cdots, |\mathcal{C}|\}$, $\widehat{p}_j = 1/|\mathcal{C}|$, for event B_j'.
- Let us call Y_{BRD} (BRD: Balanced Recursive Decomposition) the corresponding estimator. Using this uniform distribution and the likelihood ratio $p_j/\widehat{p}_j$ to keep the estimator unbiased, we formally write

$$\begin{aligned} Y_{BRD} &= q_{\mathcal{C}} + (1 - q_{\mathcal{C}}) \sum_{j=1}^{|\mathcal{C}|} \mathbf{1}_{B_j'} \frac{p_j}{\widetilde{p}_j} Y_{BRD}(\mathcal{G}_j) \\ &= q_{\mathcal{C}} + |\mathcal{C}| \sum_{j=1}^{|\mathcal{C}|} \mathbf{1}_{B_j'} \mathbb{P}[B_j] Y_{BRD}(\mathcal{G}_j). \end{aligned}$$

Results on Balanced RVR

Analyzing the relative error, we obtain

Theorem

The estimator Y_{BRD} is unbiased: $\mathbb{E}[Y_{BRD}] = Q$.
The BRD algorithm verifies the Bounded Relative Error property.

Proof: induction from the recursion; in particular, for the second claim, we start from

$$\begin{aligned}\mathbb{E}[Y_{BRD}^2] &= q_{\mathcal{C}}^2 + 2q_{\mathcal{C}}|\mathcal{C}| \left(\sum_{j=1}^{|\mathcal{C}|} \mathbb{P}[B_j]\mathbb{E}[Y_{BRD}(\mathcal{G}_j)] \right) \\ &\quad + |\mathcal{C}|^2 \left(\sum_{j=1}^{|\mathcal{C}|} (\mathbb{P}[B_j])^2 \mathbb{E}[Y_{BRD}^2(\mathcal{G}_j)] \right).\end{aligned}$$

Remarks on BRD

- Intuition behind BRD: to make sure that probability of each event B_j' is $\Theta(1)$, so that none of these events is rare under IS.
- As a consequence, the probability $\mathbb{P}(B_j)$ is squared in the likelihood ratio (which was not the case for RVR), and BRE can be obtained.
- Note that any choice of distribution such that probability of each B_j' is $\Theta(1)$, leads to BRE as well.
- For some network topologies and link unreliability values, standard Monte Carlo has lower variance than BRD. Thus, the BRD estimator does not guarantee variance reduction (implicitly, with respect to the crude estimator) in all contexts.

Outline

Intermezzo

- Move for a second to Markov chains.
- Consider a Markov chain X on some discrete state space S, with transition probability matrix (t.p.m.) P, and a cost (or reward) function c defined on the transitions; for instance, $c\colon S^2 \to \mathbb{R}_{\geq 0}$.
- We are interested in the mean cumulated cost up to some stopping time τ, $\mu = \mathbb{E}(Y)$, where $Y = \sum_{j=1}^{\tau} c(X_{j-1}, X_j)$.
- To estimate μ, suppose we use Importance Sampling (IS), with the change of measure given by the new t.p.m. P':

$$P'_{i,j} = \frac{P_{i,j}\left[c(i,j) + \mu_j\right]}{\sum_{k \in S} P_{i,k}\left[c(i,k) + \mu_k\right]},$$

where $\mu_h = \mathbb{E}(Y \mid X_0 = h)$.

- It can be proved that this change of measure is the zero-variance one (one run, exact value at the output), and it is useless in practice because we need the initial target (actually, you even need much more information than just the target) to implement it.
- "Zero-variance idea": use the new dynamics

$$P^*_{i,j} = \frac{P_{i,j}\left[c(i,j) + \mu^*_j\right]}{\sum_{k\in S} P_{i,k}\left[c(i,k) + \mu^*_k\right]},$$

  where μ^*_h is some (any) approximation to μ_h, even a poor one.
- This is a hot topic in the area; in general, this approach leads to very good (efficient) results.
- The heart of the method lies in the problem of finding an approximation μ^* to μ that can be computed very fast.
- In the paper, we use the same idea (in spite of the fact that we are in a static environment).

Zero-variance IS

- In this context, the (exact) *zero-variance change of measure* chooses the best possible IS scheme for the first working link on the cut, that is, samples B_j' with a probability $\widehat{p}_j$ where

$$\widehat{p}_j = \frac{\mathbb{P}[B_j] q(\mathcal{G}_j)}{\sum_{k=1}^{|\mathcal{C}|} \mathbb{P}[B_k] q(\mathcal{G}_k)} \tag{1}$$

- Resulting estimator:

$$Y_{ZVD} = q_{\mathcal{C}} + \left(\sum_{k=1}^{|\mathcal{C}|} \mathbb{P}[B_k] q(\mathcal{G}_k) \right) \sum_{j=1}^{|\mathcal{C}|} \mathbf{1}_{B_j'(\mathcal{G})} \frac{1}{q(\mathcal{G}_j)} Y_{ZVD}(\mathcal{G}_j).$$

Theorem

Y_{ZVD} has variance $\mathrm{Var}[Y_{ZVD}] = 0$.

Proof: induction using the recursions.

- Implementing this requires the knowledge of the $q(\mathcal{G}_i)$, but in that case, no need to simulate!
- Idea: use instead some (any) approximation $\hat{q}(\mathcal{G}_i)$ of $q(\mathcal{G}_i)$ plugged into (1). This gives a new estimator called Y_{AZVRD}:

$$Y_{AZVRD} = q_{\mathcal{C}} + \left(\sum_{k=1}^{|\mathcal{C}|} \mathbb{P}[B_k] \hat{q}(\mathcal{G}_k) \right) \sum_{j=1}^{|\mathcal{C}|} \mathbf{1}_{B'_j(\mathcal{G})} \frac{1}{\hat{q}(\mathcal{G}_j)} Y_{AZVRD}(\mathcal{G}_j).$$

Proposition

If for $1 \leq j \leq |\mathcal{C}|$, $\hat{q}(\mathcal{G}_j) = \Theta(q(\mathcal{G}_j))$ *as* $q(\mathcal{G}) \to 0$, Y_{AZVRD} *verifies the BRE property.*

Approx. 0-var. Recursive Decomposition

Define the *mincut-maxprob* approximation $\hat{q}(\mathcal{G})$ of $q(\mathcal{G})$ as the maximal probability of a mincut of graph $\mathcal{G}$ (this can be computed in polynomial time).

Proposition

With the mincut-maxprob approximation, $\hat{q}(\mathcal{G}_j) = \Theta(q(\mathcal{G}_j))$ *as* $q(\mathcal{G}) \to 0$*; therefore the BRE property is obtained.*

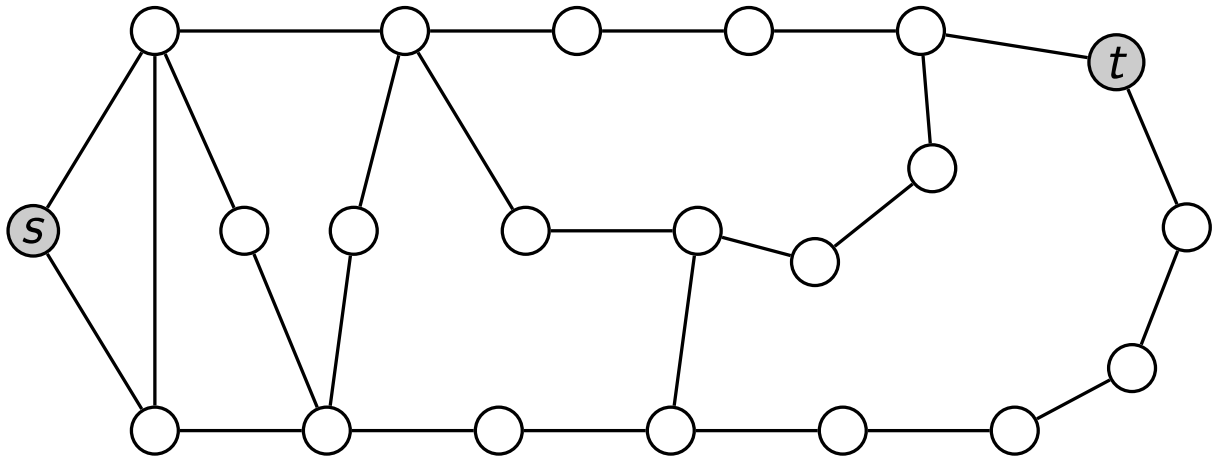
Proposition

If, $\hat{q}(\mathcal{G}_j) = q(\mathcal{G}_j) + o(q(\mathcal{G}_j))$ *as* $q(\mathcal{G}) \to 0$ *for all* $1 \leq j \leq |\mathcal{C}|$*, the Vanishing Relative Error (VRE) property (RE* $\to 0$ *as* $q(\mathcal{G}) \to 0$*, much stronger than just being bounded) is verified.*

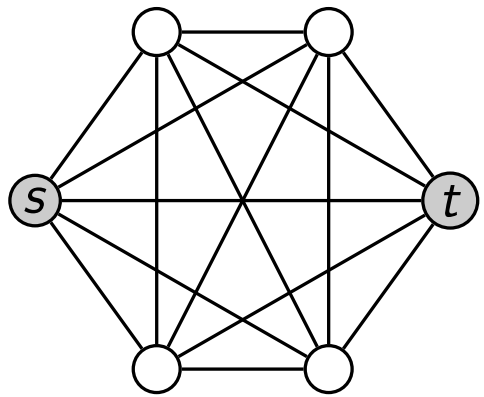
Outline

Topologies: arpanet, complete graphs,

Arpanet:

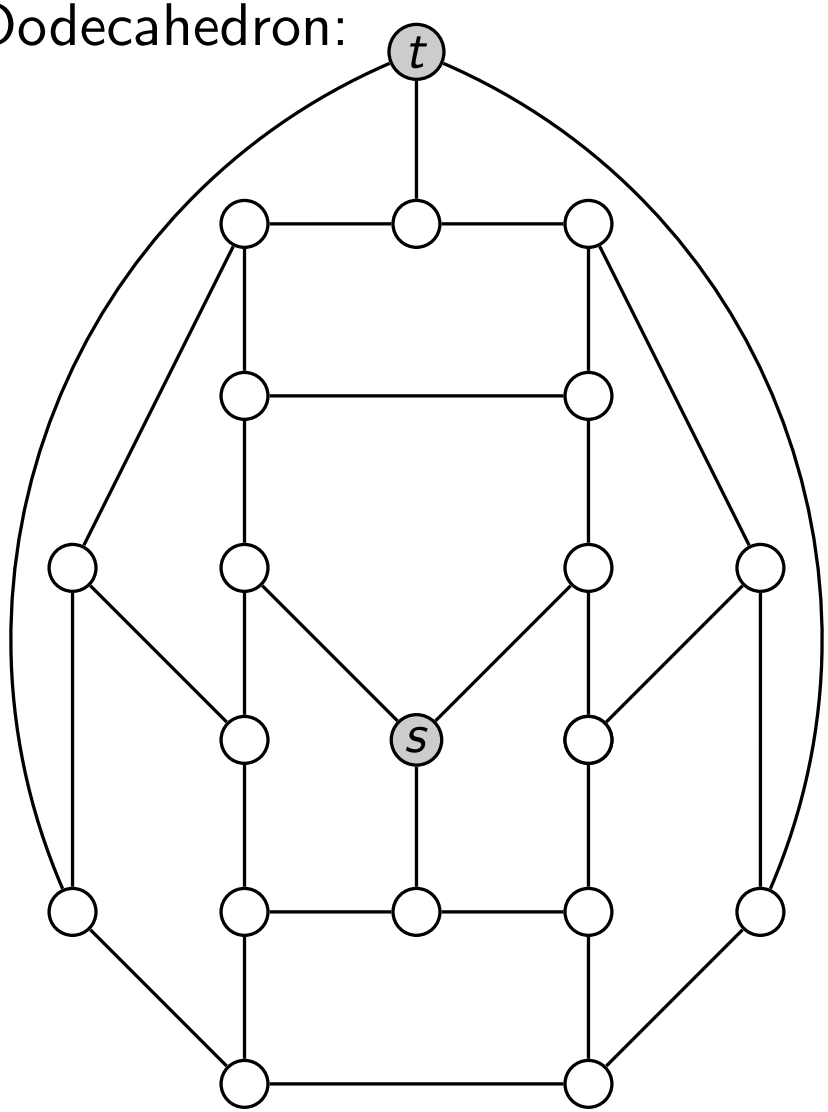


C_6:

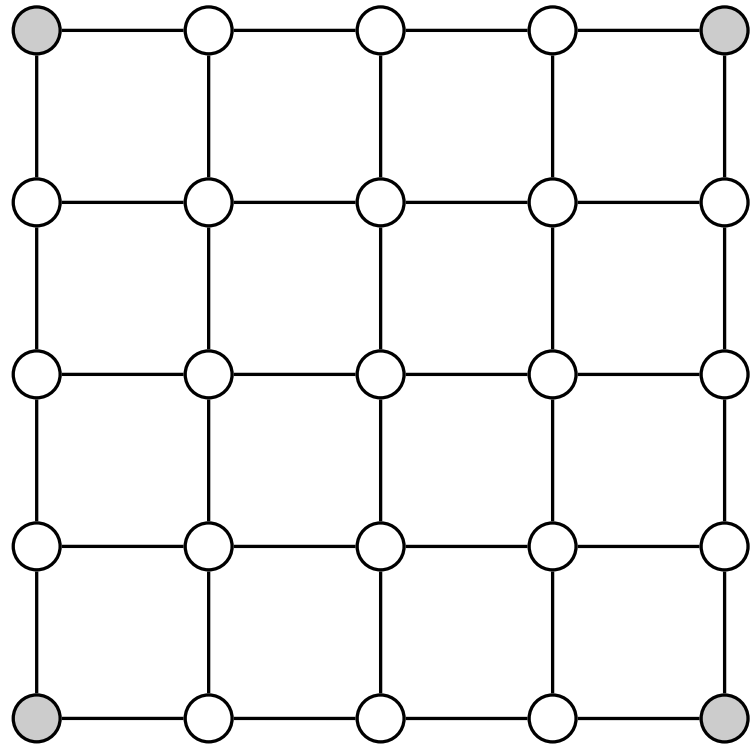


... dodecahedron and grids

Dodecahedron:



Grid, order 5:



Comparisons

The (normalized) relative error[2] for various methods and unreliabilities ε of links (homogeneous case), on the dodecahedron topology

Method	$\varepsilon = 0.1$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-4}$
[F]	2.6 e+00	1.3 e+00	4.3 e−01	1.4 e−02
[S1]	4.0 e+00	6.2 e+00	7.7 e+00	8.9 e+00
[S2]	4.6 e+00	7.1 e+00	8.6 e+00	8.8 e+00
[B]	3.0 e+00	4.2 e+00	4.3 e+00	4.4 e+00
[Z]	1.2 e+00	1.7 e−01	5.7 e−02	1.7 e−02
[R]	8.4 e−01	7.1 e−01	7.1 e−01	7.1 e−01
BRD [A]	9.5 e−01	7.0 *e*-01	7.1 e−01	7.1 e−01
AZVRD [A]	2.8 e−01	5.1 e−02	1.6 e−02	5.0 e−03

[2] Denoting RE_N the relative error, we use here $\sqrt{N} \cdot RE_N/z$, with, say, $z = 1.96$.

F G. S. Fishman. 1986. A Monte Carlo sampling plan for estimating network reliability. Operations Research 34, 4, 581–594 (method based on bounds).

S1 Z. I. Botev, P. L'Ecuyer, G. Rubino, R. Simard and B. Tuffin. 2013. Static network reliability estimation via generalized splitting. INFORMS Journal on Computing 25, 1, 56–71 (a generalization of splitting).

S2 L. Murray, H. Cancela, and G. Rubino. 2013. A splitting algorithm for network reliability. IIE Transactions 45, 2, 177–189 (another adaptation of splitting to static problems).

B I. B. Gertsbakh and Y. Shpungin. 2010. Models of Network Reliability. CRC Press, Boca Raton, FL, US (the so-called turnip method).

Z P. L'Ecuyer, G. Rubino, S. Saggadi and B. Tuffin. 2011. Approximate zero-variance importance sampling for static network reliability estimation. IEEE Transactions on Reliability 8, 4, 590–604 (another zero-variance approximation).

R H. Cancela and M. El Khadiri. 1995. A recursive variance-reduction algorithm for estimating communication-network reliability. IEEE Transactions on Reliability 44, 4, 595–602 (the original RVR technique).

A H. Cancela, M. El Khadiri, G. Rubino and B. Tuffin. 2015. Balanced and Approximate Zero-Variance Recursive Estimators for the Static Communication Network Reliability Problem. TOMACS, 25(1): 5:1–5:19 (our paper's).

Illustration of the BRE and VRE properties

Network	ε	$q(\mathcal{G})$	$\frac{\sqrt{n}\times RE_{SMC}}{c_\alpha}$	$\frac{\sqrt{n}\times RE_{RVR}}{c_\alpha}$	$\frac{\sqrt{n}\times RE_{BRD}}{c_\alpha}$	$\frac{\sqrt{n}\times RE_{AZVRD}}{c_\alpha}$
Arpanet	5 e−01	9.6398994 e−01	1.93 e−01	6.33 e−02	4.16 e−01	4.27 e−01
Arpanet	3 e−01	6.8150724 e−01	6.84 e−01	3.20 e−01	1.10 e+00	1.35 e+00
Arpanet	1 e−01	9.5422918 e−02	3.08 e+00	1.27 e+00	2.01 e+00	3.24 e+00
Arpanet	1 e−03	6.0558106 e−06	4.06 e+02	2.09 e+01	1.24 e+00	9.67 e−01
Arpanet	1 e−05	6.0005600 e−10	4.08 e+04	2.11 e+02	1.26 e+00	9.82 e−02
Dod	5 e−01	7.0974499 e−01	6.39 e−01	1.77 e−01	9.17 e−01	5.17 e−01
Dod	3 e−01	1.6851806 e−01	2.22 e+00	5.70 e−01	1.93 e+00	7.70 e−01
Dod	1 e−01	2.8796013 e−03	1.86 e+01	8.37 e−01	9.53 e−01	2.76 e−01
Dod	1 e−03	2.0060181 e−09	2.23 e+04	7.08 e−01	7.06 e−01	1.59 e−02
Dod	1 e−05	2.0000600 e−15	2.24 e+07	7.07 e−01	7.07 e−01	1.58 e−03
Grid5	5 e−01	9.6062484 e−01	2.02 e−01	2.66 e−02	4.55 e−01	1.01 e−01
Grid5	3 e−01	5.2094890 e−01	9.59 e−01	1.53 e−01	1.17 e+00	2.29 e−01
Grid5	1 e−01	4.8160510 e−02	4.45 e+00	1.40 e−01	1.09 e+00	1.35 e−01
Grid5	1 e−03	4.0080020 e−06	4.99 e+02	1.58 e−02	1.14 e+00	1.37 e−02
Grid5	1 e−05	4.0000800 e−10	5.00 e+04	1.58 e−03	1.15 e+00	1.37 e−03
C6	5 e−01	7.6416016 e−02	3.48 e+00	1.15 e−01	3.43 e−01	1.12 e−01
C6	3 e−01	5.2672775 e−03	1.37 e+01	9.61 e−02	5.32 e−01	9.06 e−02
C6	1 e−01	2.0076587 e−05	2.23 e+02	1.78 e−02	7.53 e−01	1.71 e−02
C6	1 e−03	2.0000000 e−15	2.24 e+07	1.58 e−05	8.65 e−01	1.58 e−05
C6	1 e−05	2.0000001 e−25	$2.24e+12$	1.89 e−08	8.66 e−01	1.89 e−08
C10	5 e−01	1.9550825 e−02	7.08 e+00	2.10 e−01	3.65 e+01	3.13 e−01
C10	3 e−01	1.9690832 e−04	7.13 e+01	2.21 e−01	7.33 e+01	4.35 e−01
C10	1 e−01	1.0000004 e−08	1.00 e+04	3.33 e−01	1.04 e+02	5.95 e−01
C10	1 e−03	5.9991786 e−27	$1.29e+13$	5.27 e+00	1.17 e+01	4.99 e−01
C10	1 e−05	4.1102231 e−45	$1.56e+22$	7.69 e+01	2.63 e+00	2.70 e−01

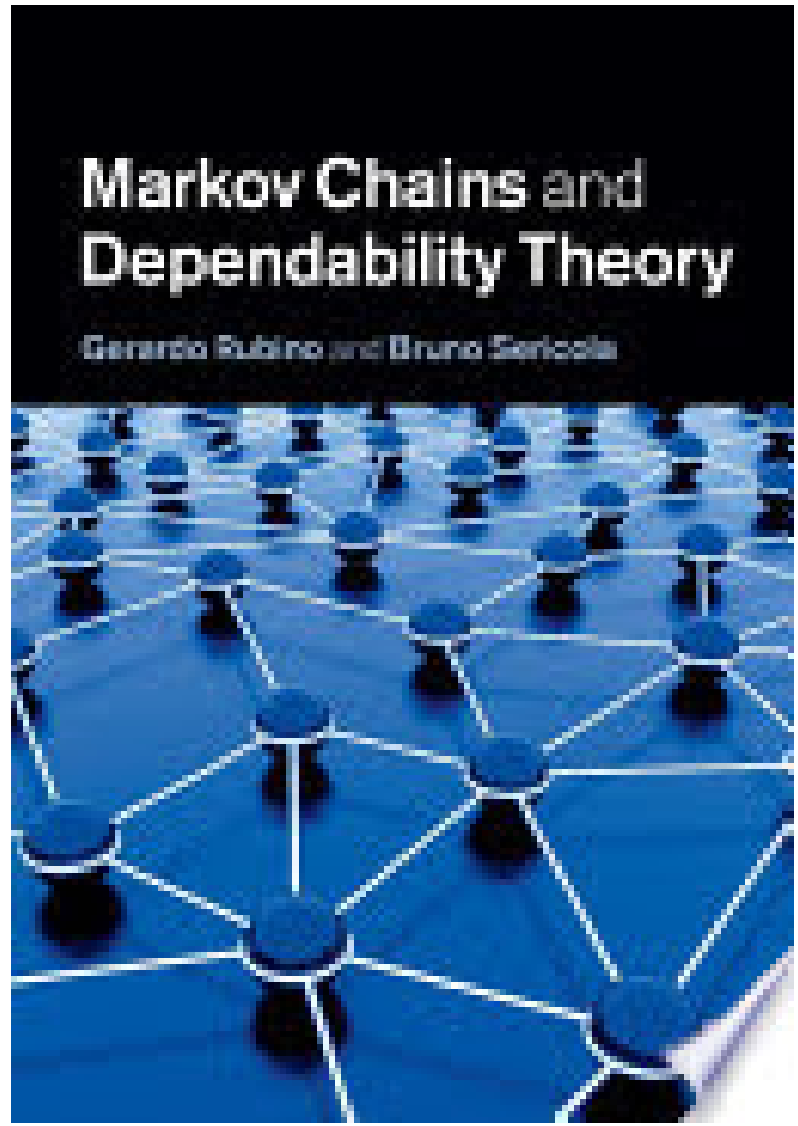
Outline

Conclusions

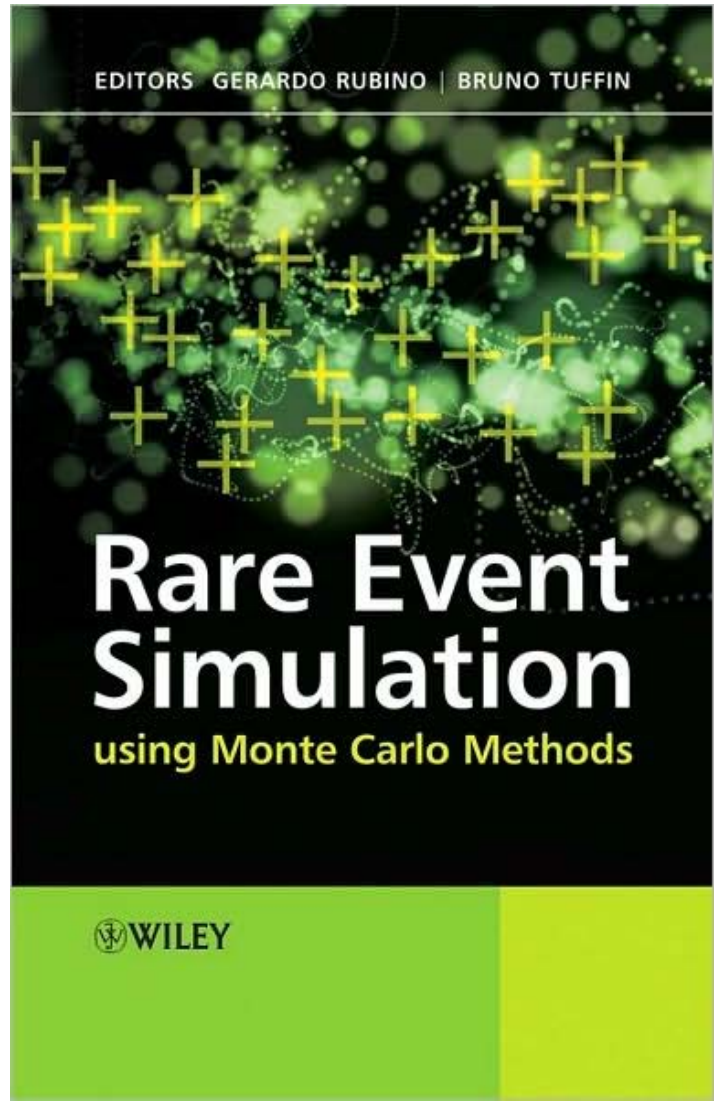
- When we must estimate the probability of a rare event defined on a complex model precluding any exact analysis, straightforward (naive, crude, standard) simulation doesn't work.
- This is also the case if we want to analyze some variable that (partially) depends on a rare event.
- Good news: the three families of methods previously mentioned may provide procedures that can evaluate in a very short time variables that would take excessive computing time with standard approaches.
- Bad news: there is no universal approach, and, often, ad hoc methods must be developed, with some effort. In difficult situation, finding an efficient technique becomes a research problem.
- But, all in all, very fast procedure can be exhibited for a large number of applications.

Recent book (Cambridge University Press 2014)

- Include a chapter on bounding techniques
- and a chapter on rare transient events (evaluation of the reliability at time t, $R(t)$, when t is small enough such that $R(t)$ is very close to one.



- Book edited by G. Rubino and B. Tuffin. Covering many fundamental issues and techniques in the area.
- Content:
  - *Introduction to Rare Event Simulation* (G. Rubino and B.Tuffin)
  - Part I: Theory
    - *Importance Sampling* (P. L'Ecuyer, M. Mandjes and B. Tuffin)
    - *Splitting Techniques* (P. L'Ecuyer, F. Le Gland, P. Lezaud and B. Tuffin)
    - *Robustness Properties and Confidence Interval Reliability Issues* (P. Glynn, G. Rubino and B. Tuffin)



2009, Wiley

Book (cont'd)

The rest of the chapters (Part II: Applications)

- *Rare Event Simulation for Queues* (J. Blanchet and M. Mandjes)
- *Markovian Models for Dependability Analysis* (G. Rubino and B. Tuffin)
- *Rare Event Analysis by Monte Carlo Techniques in Static Models*, (H. Cancela, M. El Khadiri and G. Rubino)
- *Rare Event Simulation and Counting Problems* (J. Blanchet and D. Rudoy)
- *Rare Event Estimation for a Large-Scale Stochastic Hybrid System with Air Traffic Application* (H. A. P. Blom, G. J. Bakker and J. Krystul)
- *Particle Transport Applications* (T. Booth)
- *Rare Event Simulation Methodologies in Systems Biology* (W. Sandmann)

Thanks.